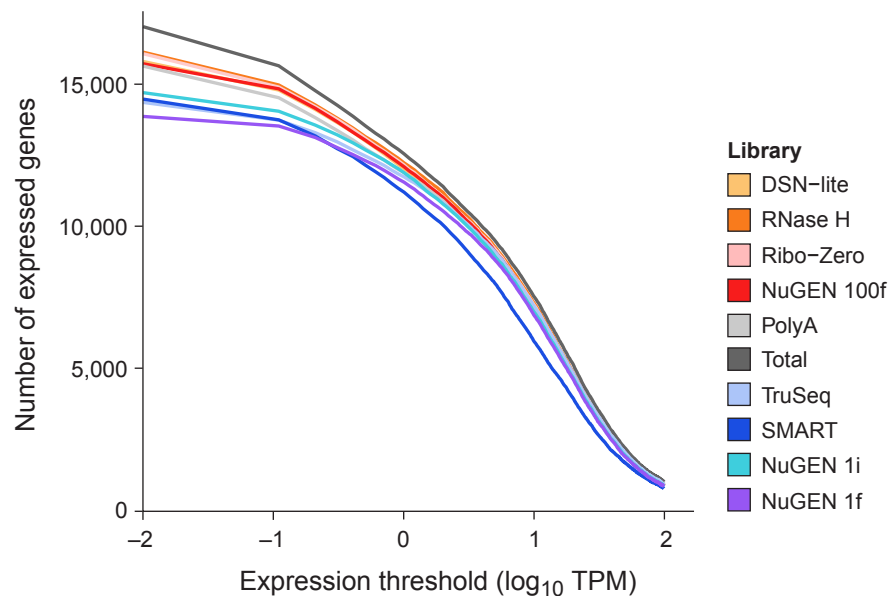


Comprehensive comparative analysis of RNA sequencing methods for degraded or low input samples

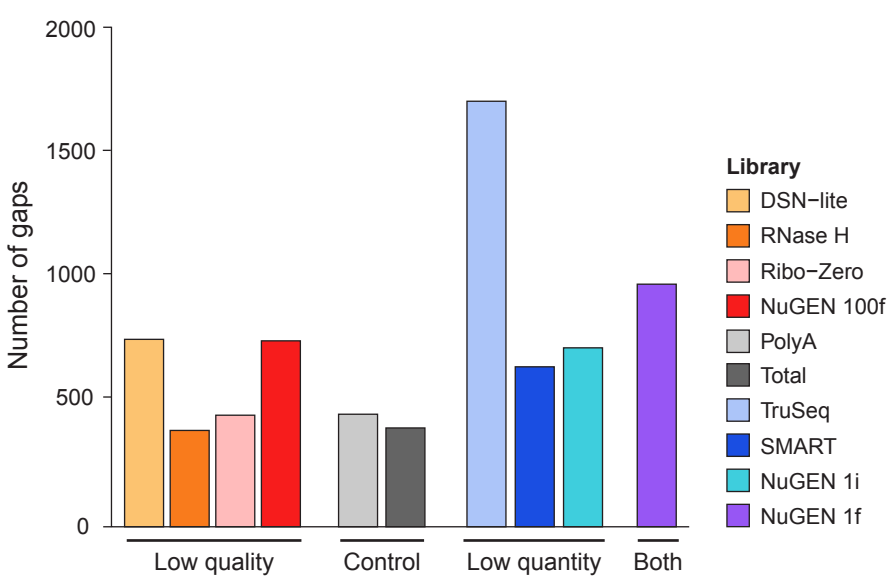
Xian Adiconis, Diego Borges-Rivera, Rahul Satija, David S. DeLuca, Michele A. Busby, Aaron M. Berlin, Andrey Sivachenko, Dawn Anne Thompson, Alec Wysoker, Timothy Fennell, Andreas Gnirke, Nathalie Pochet, Aviv Regev, Joshua Z. Levin

Suppl. Figure 1.
Number of genes detected.



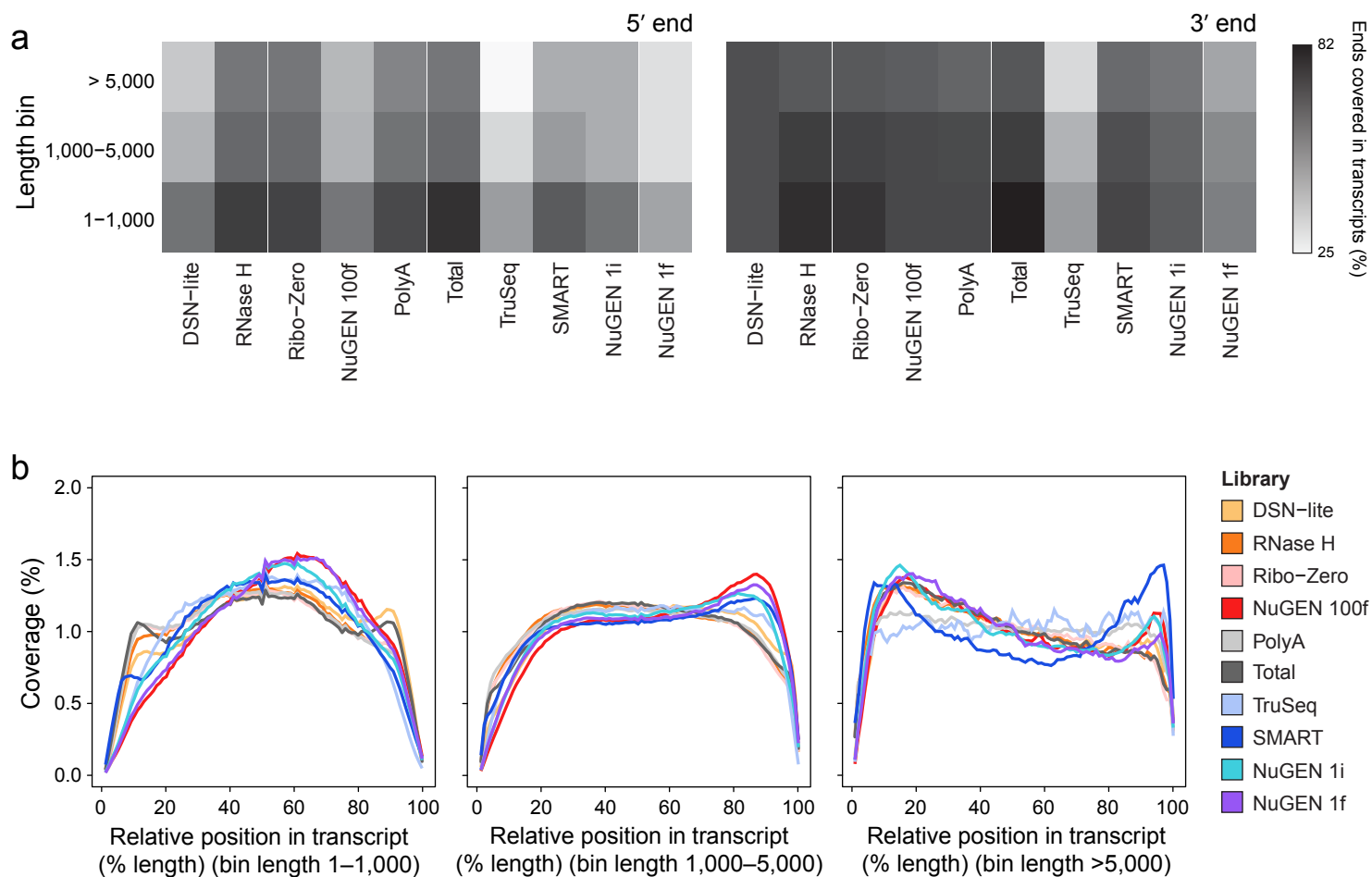
For each library, shown is the number of genes (y-axis) (from the set of all genes expressed in the control Total library) expressed at a level higher than a given TPM threshold (x-axis). The relative rankings are largely robust to changes in the threshold, with at most small flips and infrequent changes in order. The most critical comparison of SMART and NuGEN 1i is stable across these thresholds.

Suppl. Figure 2.
Continuity of coverage.



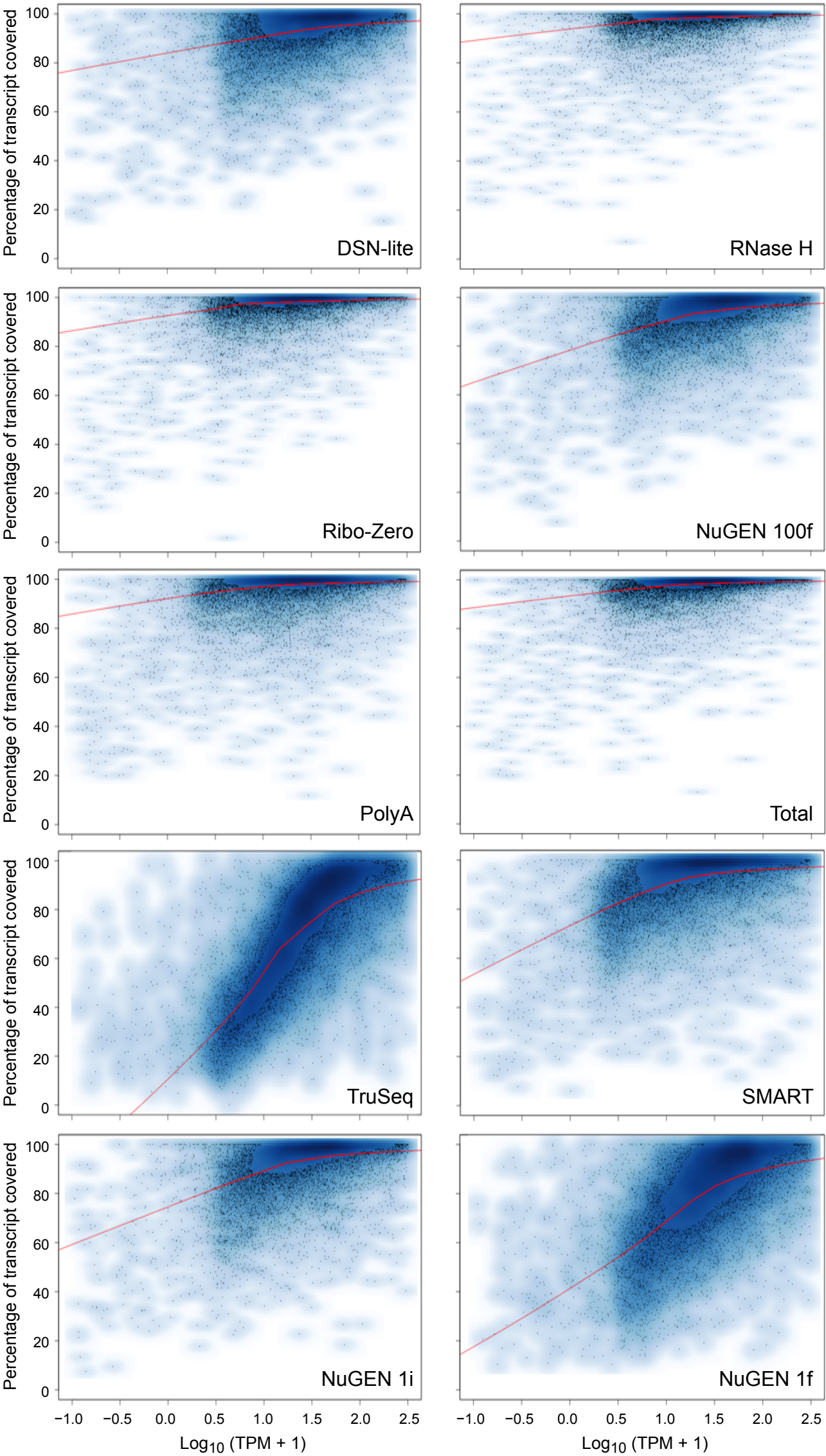
For each library, shown is the number of gaps for the top 1,000 expressed transcripts.

Suppl. Figure 3.
Length effects on 5' and 3' coverage.



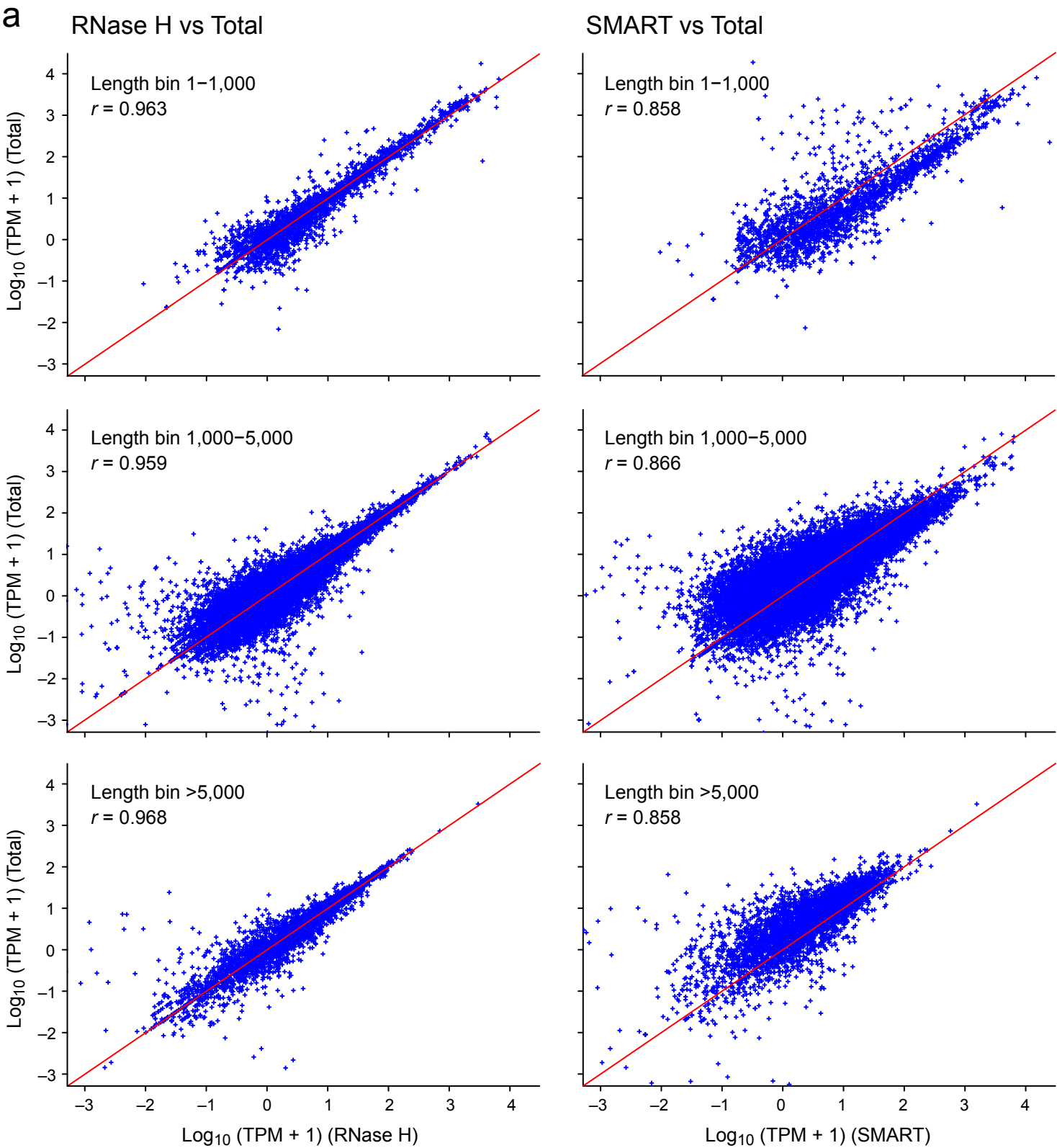
(a) Effect on coverage of 5' and 3' ends. Shown are the percent of 5' (left) and 3' (right) ends (color scale, far right) in each library (columns) for transcripts with different lengths (rows). **(b)** Effect on normalized coverage by position. For each library, shown is the relative coverage (y-axis) at each relative position along the transcripts' length for short (left), medium (middle) and long (right) transcripts.

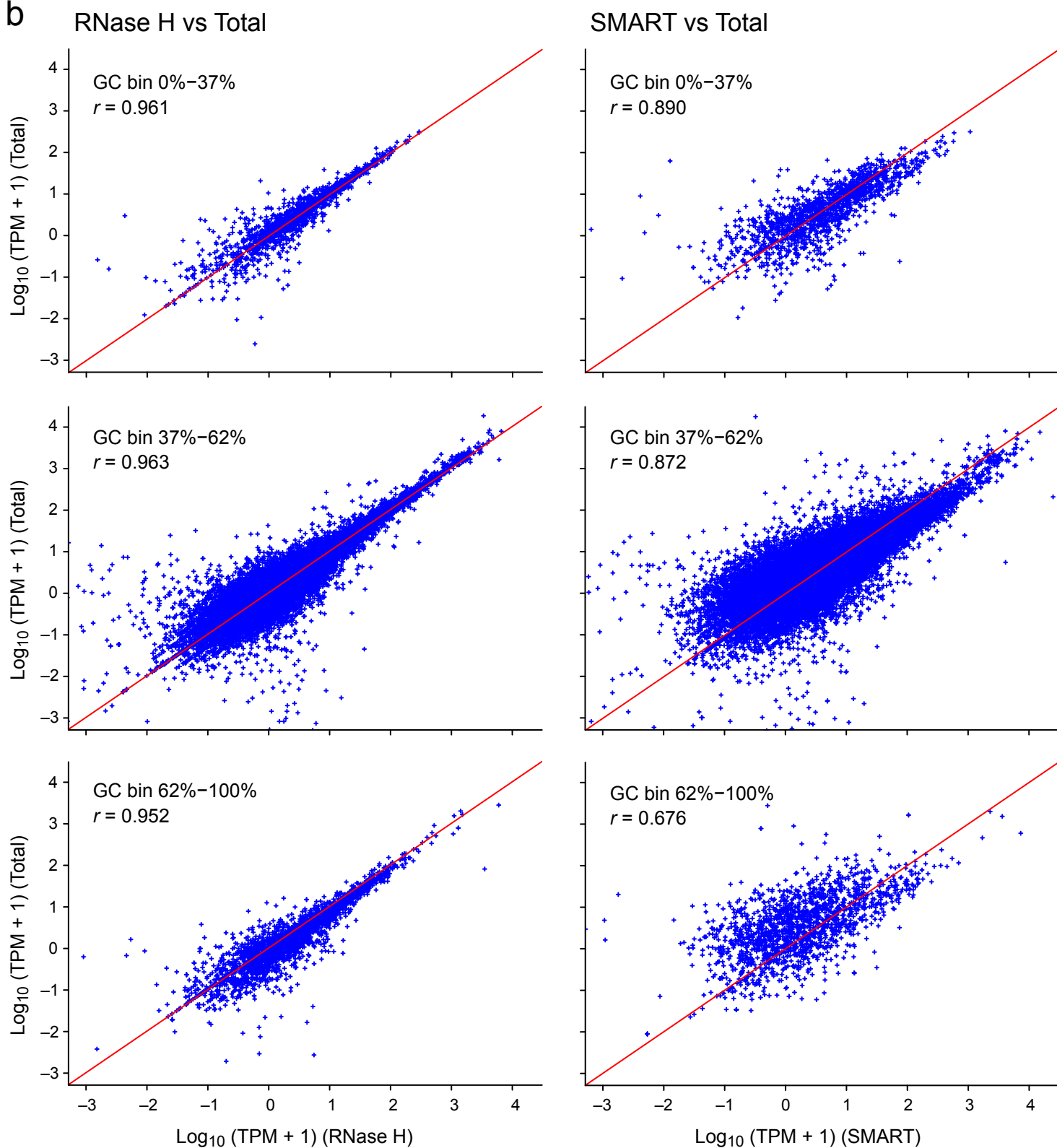
Suppl. Figure 4.
Percentage of gene covered at different expression levels.



For each library, shown is the proportion of each transcript covered by reads (y-axis, blue dots) at each expression level (x-axis), as well as the Lowess fits of this data (red curve, also shown in **Fig. 2e**). We use a density diagram to indicate the number of genes in each portion of the plot.

Suppl. Figure 5.
Effect of length and GC content on expression measures.

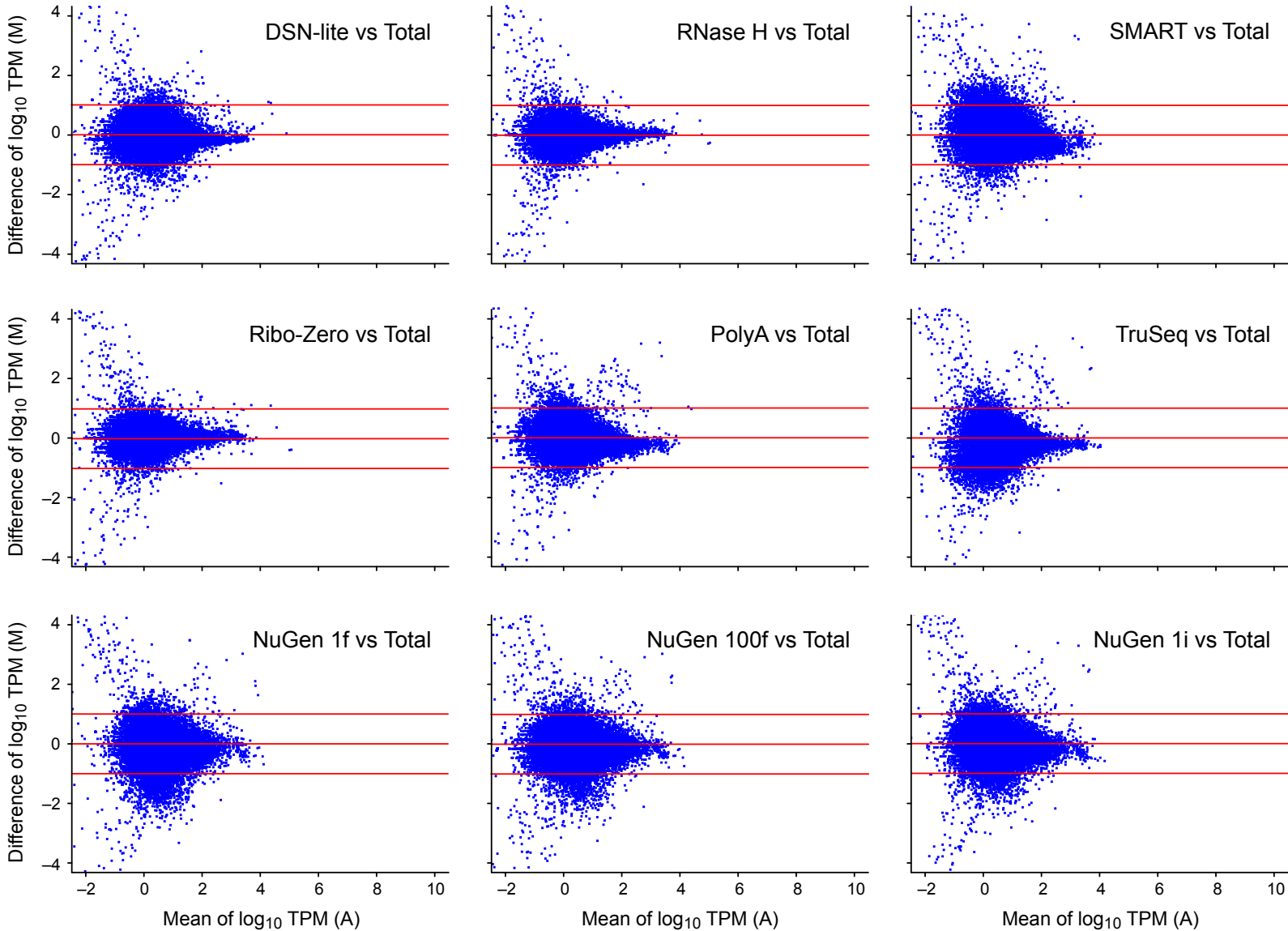


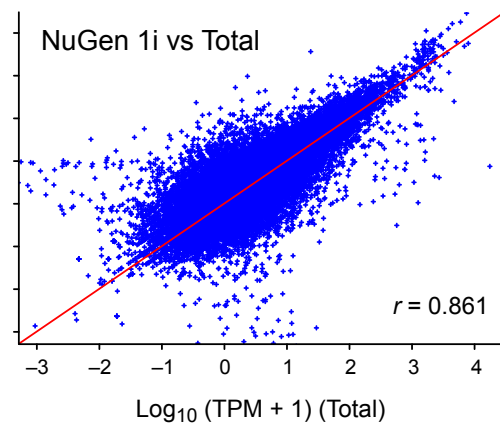
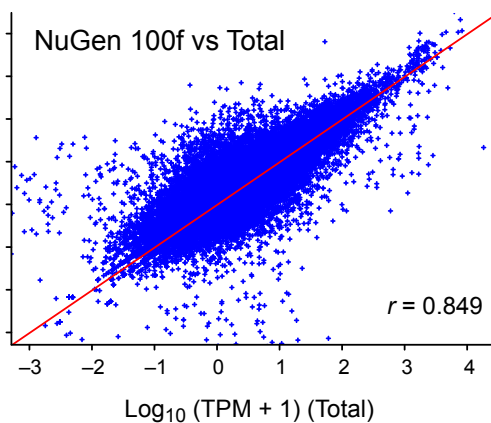
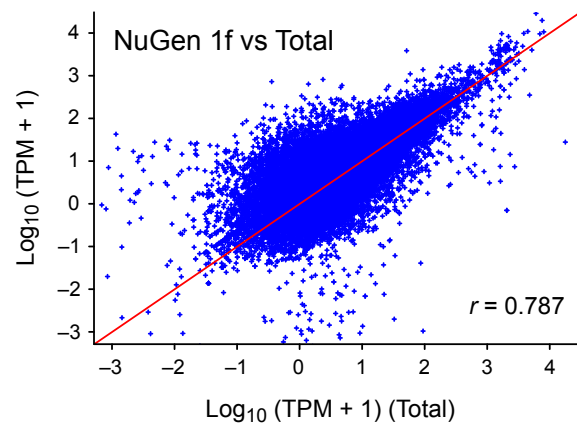
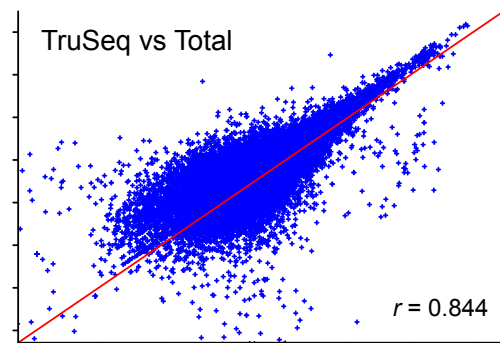
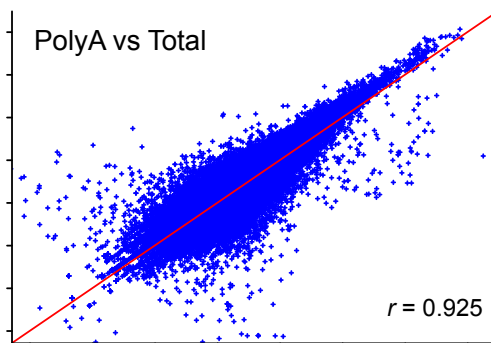
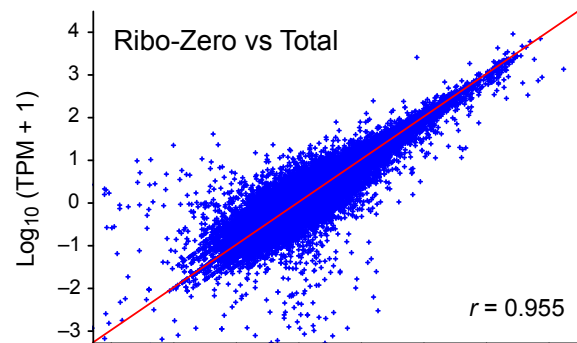
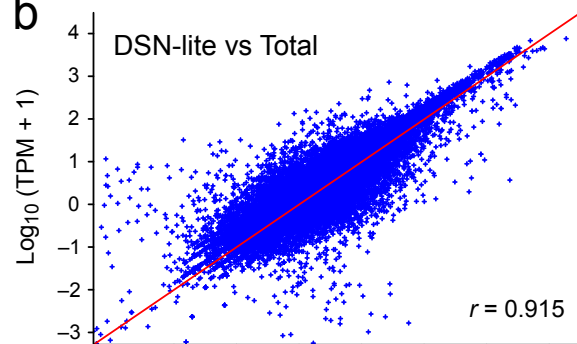
b

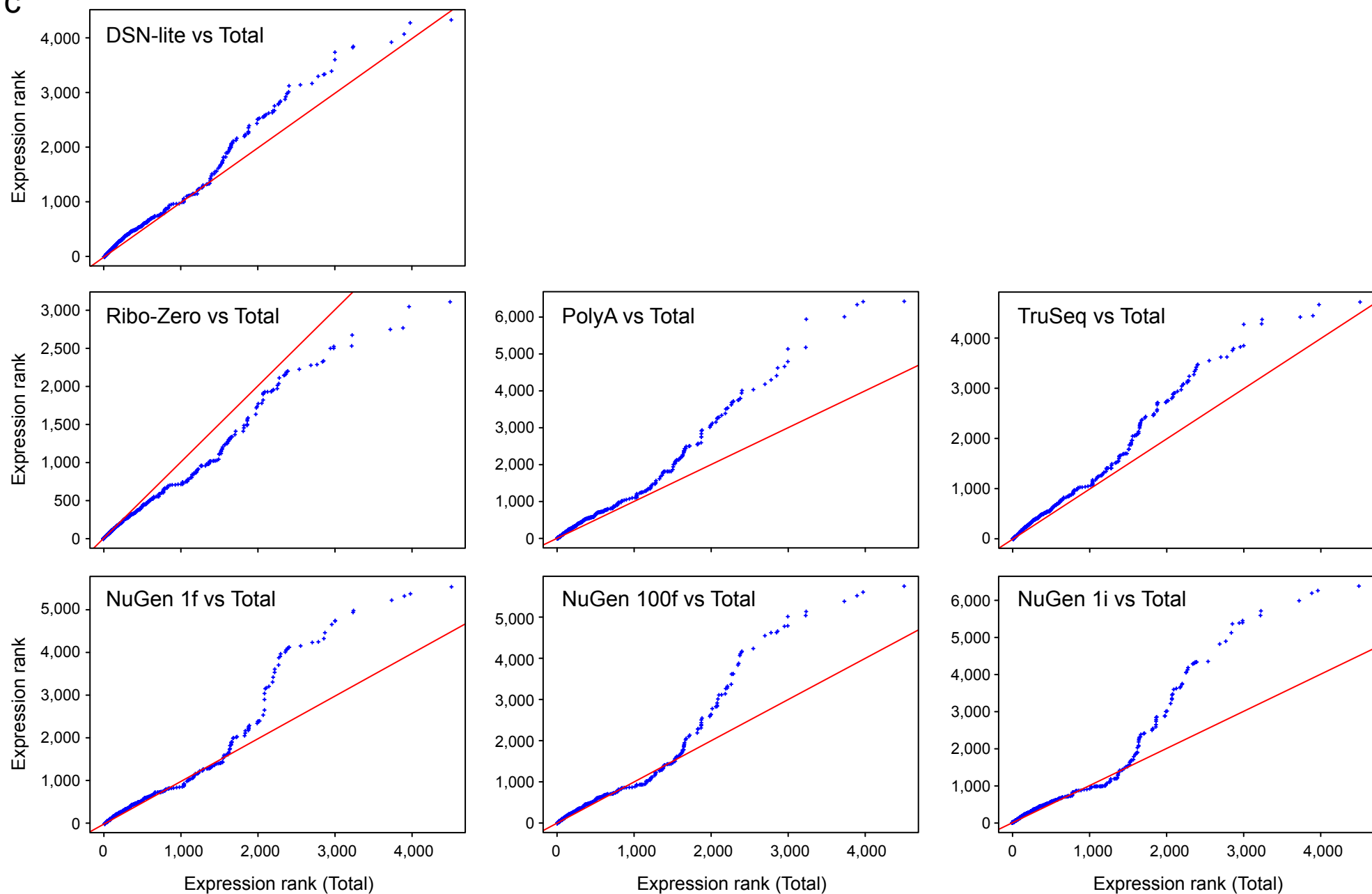
Shown are illustrative scatter plots between a low-quality library (RNase H, left, y-axis) or a low-quantity library (SMART, right, y-axis) and the control Total library (x-axis) when transcripts are binned based on (a) length or (b) GC content. Pearson correlation with the control Total library is included in the upper left hand corner of each plot.

Suppl. Figure 6.
MA, scatter, and Q-Q expression plots.

a



b

C

(a) MA plots. For each library shown is the difference of each transcript in log expression levels from the control Total library ('M', y-axis) versus that transcript's average expression in the given library and the control library ('A', x-axis). The closer the points are to the $y = 0$ line, the more similar the samples; we added $y = 1$ and $y = -1$ lines for reference. (b) Scatter plots. For each library, shown is the log expression level of the transcript versus its level in the control Total library. The Pearson correlation coefficient with the control Total library is included in the lower right hand corner of each plot. (c) Q-Q plots. For each library, transcript expression levels are plotted by increasing ranks ('quantiles', y-axis) versus the similarly-ordered expression levels in the control library (x-axis).